Time Series Homework(附 R 代码实现)

杨宸宇　　2016301550186

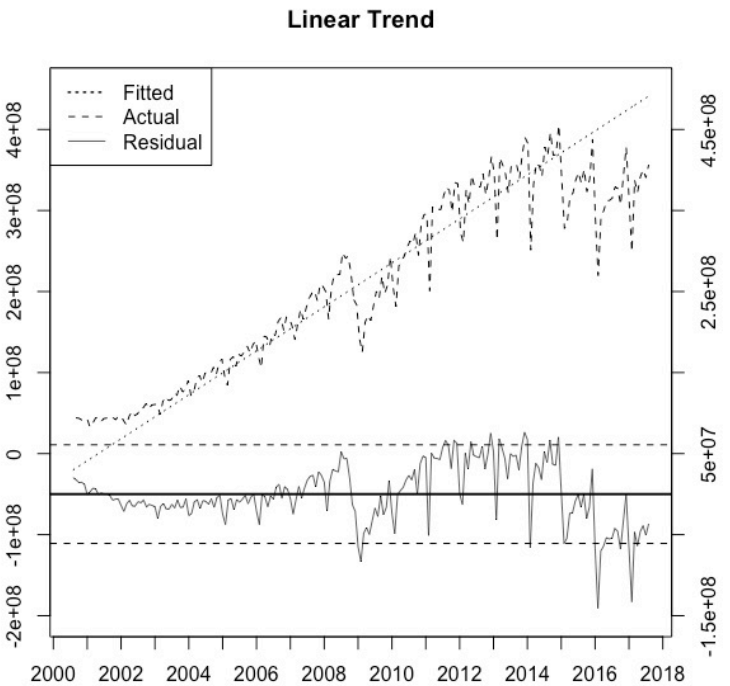
1.1 Trends

Initially, I modify the dataframe into the standard way, and I try to use different models to judge whether there are trends in the data.

Firstly, I use linear model to fit in the row data, and the results and plot show as followings:

MES = 1.349423e+15
AIC = 7729.648

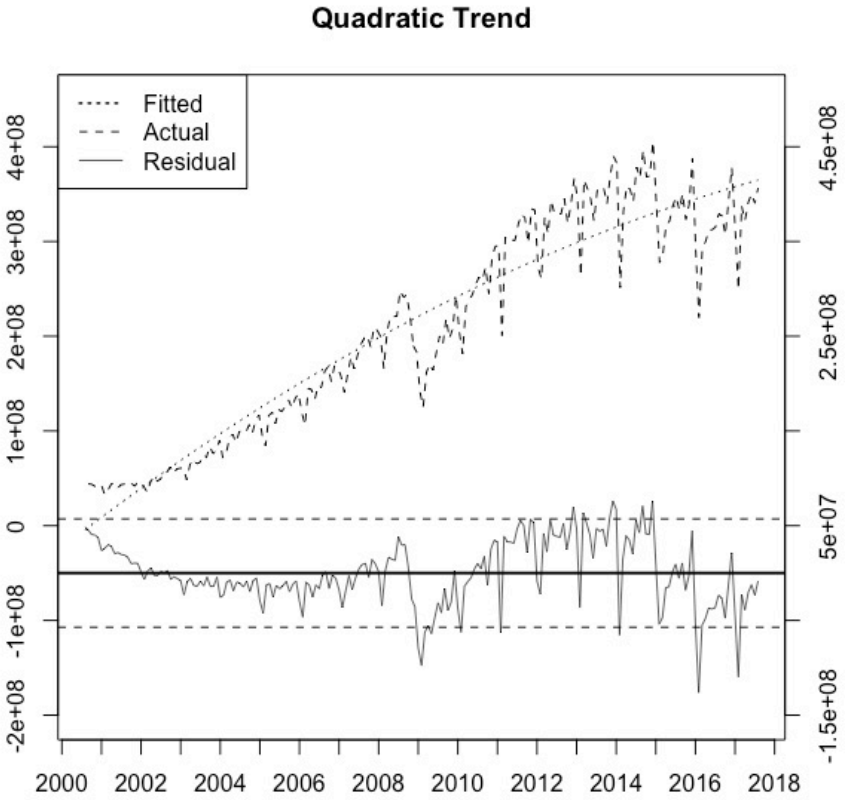
```
Call:
lm(formula = Y ~ X1)

Residuals:
       Min         1Q     Median         3Q        Max
-140714455  -17650223   -3947734   23280287   76477474

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 22049785    5175436    4.26 3.12e-05 ***
X1           1809421      43568   41.53  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36920000 on 203 degrees of freedom
Multiple R-squared:  0.8947,    Adjusted R-squared:  0.8942
F-statistic:  1725 on 1 and 203 DF,  p-value: < 2.2e-16
```



Then I use the quadratic model to fit in the row data, and the results and plot show as followings:

MSE = 1.18585e+15
AIC = 7705.158

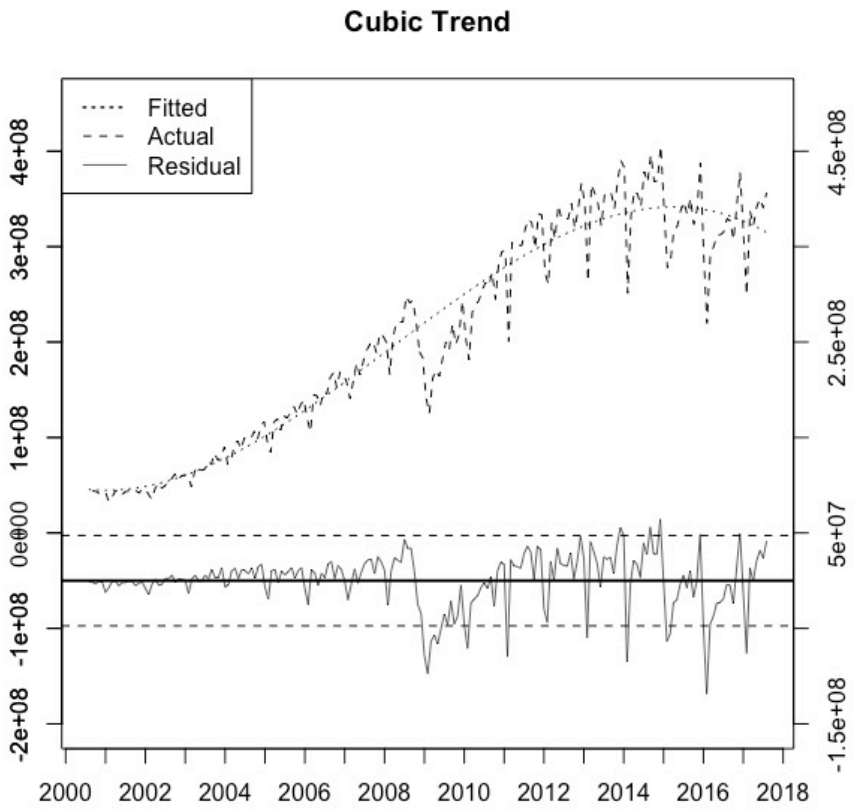
```
Call:
lm(formula = Y ~ X1 + X2)

Residuals:
       Min         1Q     Median         3Q        Max
-126202428  -18192205   -4326437   24666448   76042086

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6970184.7  7340258.1  -0.950    0.343
X1           2650579.3   164529.0  16.110  < 2e-16 ***
X2             -4083.3      773.6  -5.279 3.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34690000 on 202 degrees of freedom
Multiple R-squared:  0.9075,    Adjusted R-squared:  0.9065
F-statistic: 990.5 on 2 and 202 DF,  p-value: < 2.2e-16
```



After that, I use the cubic model and even more complex model to fit in the row data, and the results and plot for cubic model are followings:

MSE = 8.094123e+14
AIC = 7645.482

```
Call:
lm(formula = Y ~ X1 + X2 + X3)

Residuals:
       Min         1Q     Median         3Q        Max
-118667487   -7285952    2067613   16988582   64333192

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.589e+07  8.176e+06   5.612 6.56e-08 ***
X1          -3.916e+05  3.429e+05  -1.142    0.255
X2           3.275e+04  3.863e+03   8.477 4.92e-15 ***
X3          -1.192e+02  1.233e+01  -9.669  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28730000 on 201 degrees of freedom
Multiple R-squared:  0.9368,    Adjusted R-squared:  0.9359
F-statistic: 993.8 on 3 and 201 DF,  p-value: < 2.2e-16
```

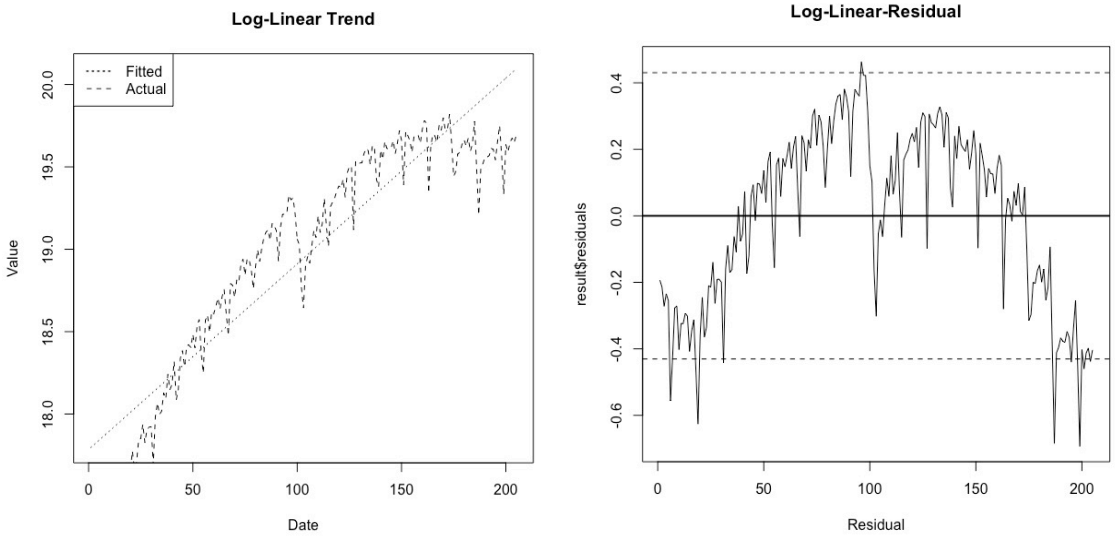


Finally, I try the log model, and I have found the results and plot for it, and them show as followings:

MSE = 0.06752925

AIC = 35.24995

```
Call:
lm(formula = Y ~ X1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.69327 -0.21059  0.06689  0.21041  0.46272

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.778e+01  3.661e-02  485.76   <2e-16 ***
X1          1.127e-02  3.082e-04   36.57   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2611 on 203 degrees of freedom
Multiple R-squared:  0.8682,    Adjusted R-squared:  0.8676
F-statistic:  1337 on 1 and 203 DF,  p-value: < 2.2e-16
```



In the end, by analysis all these models, I find the row data truly has trend, and the 4th square can fit it well. In another word, I use this model to take apart the trend in the further process.

1.2 Seasonality

In this question, I set 12 dummy variables to judge whether there are seasonal trend in the row data, and the results are as followings:

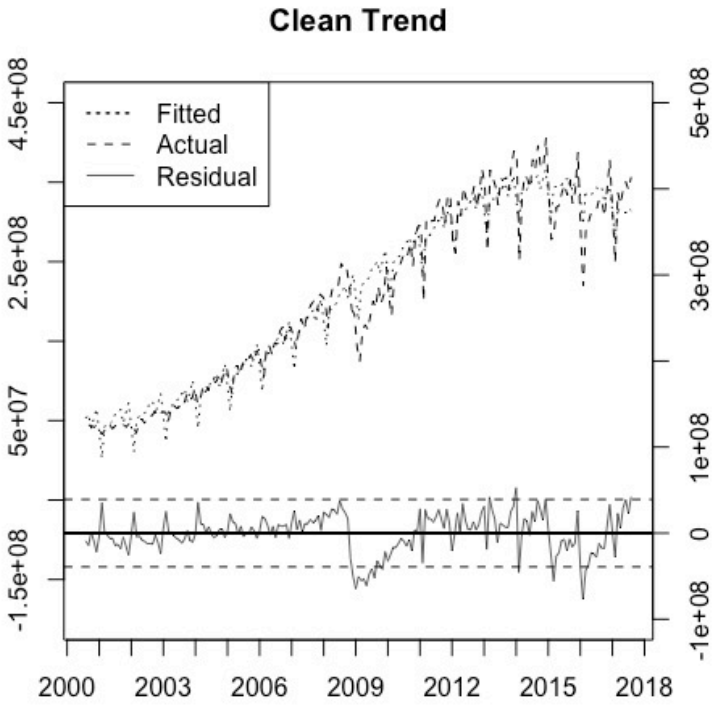
MSE = 5.106391e+14
AIC = 7558.435

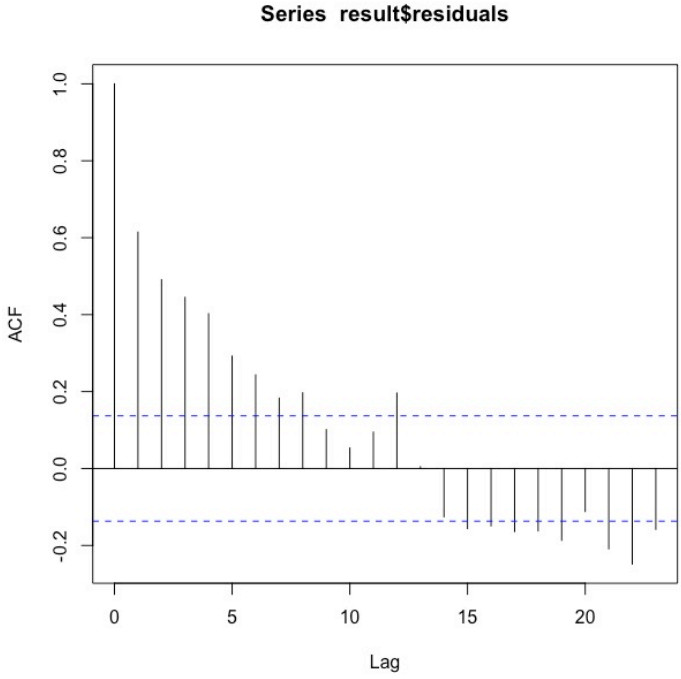
```
Call:
lm(formula = Y ~ 0 + X_time1 + X_time2 + X_time3 + X_time4 +
    X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 +
    X12)

Residuals:
      Min        1Q    Median        3Q       Max
-72344420 -11721017   1335743  14679787  51642361

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
X_time1  9.549e+05  5.677e+05   1.682  0.09422 .
X_time2  3.901e+03  1.116e+04   0.349  0.72716
X_time3  9.679e+01  8.131e+01   1.190  0.23541
X_time4 -5.221e-01  1.958e-01  -2.666  0.00834 **
X1       2.179e+07  1.011e+07   2.156  0.03233 *
X2      -1.740e+07  1.013e+07  -1.717  0.08760 .
X3       2.754e+07  1.016e+07   2.712  0.00731 **
X4       3.147e+07  1.018e+07   3.092  0.00229 **
X5       3.003e+07  1.020e+07   2.944  0.00365 **
X6       3.267e+07  1.022e+07   3.198  0.00162 **
X7       4.112e+07  1.023e+07   4.019 8.43e-05 ***
X8       4.131e+07  9.786e+06   4.221 3.77e-05 ***
X9       4.574e+07  9.968e+06   4.589 8.10e-06 ***
X10      2.773e+07  1.001e+07   2.771  0.00615 **
X11      3.993e+07  1.004e+07   3.976 9.98e-05 ***
X12      5.183e+07  1.008e+07   5.144 6.69e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23530000 on 189 degrees of freedom
Multiple R-squared:  0.9909,    Adjusted R-squared:  0.9902
F-statistic:  1289 on 16 and 189 DF,  p-value: < 2.2e-16
```

Through the results we can easily find there are great seasonal trends in the data(except February), especially in July, August, September, November and December. And we can draw a conclusion that the seasonality truly exists.

1.3 Analysis and plot data without trends and seasonality

Cleaned data plot & Residual ACF plot:





Through the plot, we can see the trend is departed most after my operations on it. However, if we focus on the ACF plot, we can find that there may be still some dependence structures in the cleaned data.

附录

第一题的 R 代码实现

```
# Topic: The Second Homework in Analysis of Time Series

# Name: YANG CHENYU
# Class: Financial Engineering 2
# Student ID:2016301550186

# load package-------------------------------------------------
library(tseries)
library(lmtest)
library(tidyverse)

# set workspace------------------------------------------------
setwd('/Users/mac/Desktop/R_Time_Analysis')


# input data---------------------------------------------------
data <- read.csv('hw2_data.csv', head = T, sep = ',')
row=nrow(data)

# prepare and integrate date-----------------------------------------------
data <- data %>%
  # divide dates into year & month
  separate(Year, into = c("Year", "Mon"), sep = '/')
data$Year <- as.double(data$Year)
data$Mon <- as.double(data$Mon)

# let the Month column maintain double and change Date into character
data["Month"] <- data$Mon
# add day as 1 in order to change Date from character to date(double in R)
data["add"] = 1
data <- data %>%
  arrange(Year, Mon) %>%
  unite(Date,c("Year", "Mon", "add"), sep = '-')
data$Date = as.Date(data$Date)
data = data[c(1,3,2)]

# add the dummy variables for each month
data <- cbind(data[1:2],1,2,3,4,5,6,7,8,9,10,11,12,data[3])
data$index = 0
data[3:14] = 0
data = data[c(16,1:15)]

for(i in seq_along(data$Month)){
  data[i, 1] = i
  a = data[i, 3]%%12
  if(a == 0){
    data[i, 15] = 1
```

```
  }
  else{
    data[i, 3 + a] = 1
  }
}
# have a simple look at data
plot(x = data$Date, y = data$Value, xlab = 'Date', ylab = 'Value', main = 'First
Look')


# regression analysis-----------------------------------------------


Y = data$Value

# Problem_1
# estimate the trends in the dataset


# 1.Linear Trend
X1 = data$index
result=lm(Y~X1)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-1)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))


# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# plot
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-
120000000,400000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

par(new = TRUE)
plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-
200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-
200000000,450000000,by=100000000)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Linear Trend', ylim = c(-
150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes =
F)
abline(h = 0, lwd = 2)
```

```
abline(h = up, lty= 2)
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-
150000000,500000000,by=100000000)))
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,
0.7))
box()


# 2. Quadratic Trend
X1 = data$index
X2 = X**2
result=lm(Y~X1+X2)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-2)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))


# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# plot
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-
200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)
axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)))

par(new = TRUE)
plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-
200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-
200000000,450000000,by=100000000)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Quadratic Trend', ylim = c(-
150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes =
F)
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-
150000000,500000000,by=100000000)))
```

```
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,
0.7))
box()

# 3.Cubic Trend
X1 = data$index
X2 = X1**2
X3 = X1**3
result=lm(Y~X1+X2+X3)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-3)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))


# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down


# plot
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-
200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)
axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)))

par(new = TRUE)
plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-
200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-
200000000,450000000,by=100000000)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Cubic Trend', ylim = c(-
150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes =
F)
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-
150000000,500000000,by=100000000)))
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,
0.7))
```

```r
box()



# 4.Log-Linear Trend
Y = log(data$Value)
X1 = data$index
result=lm(Y~X1)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-1)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))



# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# plot
plot.new()
plot(x = data$index, y = result$fitted.values, main = 'Log-Linear Trend', type =
'l', xlab = 'Date', ylab = 'Value', lty = 3)
lines(x = data$index, y = Y, lty = 2, lwd = 1)
legend('topleft', c('Fitted', 'Actual'), lty=c(3, 2), lwd = c(1.5,1))

plot(x = data$index, result$residuals, type = 'l', xlab = 'Residual', main = 'Log-
Linear-Residual')
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)

# plot in one pic, but it doesn't seem good
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-1,
21), xlab = '', ylab = '', lty = 3, ldw = 1.5)
plot(x = data$index, y = Y, xlim = c(0,205), ylim = c(-1, 21),type = 'l', lty = 2,
lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-1,21,by=5)), lab = c(seq(-1,21,by=5)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Log-Linear Trend', ylim = c(-
1,21),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
```

```r
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-1,1,by=0.1)), lab = c(seq(-1,1,by=0.1)))
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,
0.7))


# Problem_2
X1 = data$`1`
X2 = data$`2`
X3 = data$`3`
X4 = data$`4`
X5 = data$`5`
X6 = data$`6`
X7 = data$`7`
X8 = data$`8`
X9 = data$`9`
X10 = data$`10`
X11 = data$`11`
X12 = data$`12`
Y = data$Value
result = lm(Y~0+X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)
summary(result)
sum(result$residuals^2)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-12)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))


# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# plot
plot.new()
plot(x = data$Date, y = result$fitted.values, main = 'Seasonality', type = 'l',
ylim = c(-190000000,400000000), xlab = 'Date', ylab = 'Value', lty = 3, ldw = 1.5)
lines(x = data$Date, y = data$Value, lty = 2, lwd = 1)
lines(x = data$Date, y = result$residuals, lty = 1, lwd = 0.7)
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,
0.7))
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)
```

```r
# Problem_3
X_time1 = data$index
X_time2 = X_time1**2
X_time3 = X_time1**3
X_time4 = X_time1**4
X1 = data$`1`
X2 = data$`2`
X3 = data$`3`
X4 = data$`4`
X5 = data$`5`
X6 = data$`6`
X7 = data$`7`
X8 = data$`8`
X9 = data$`9`
X10 = data$`10`
X11 = data$`11`
X12 = data$`12`
Y = data$Value
result =
lm(Y~0+X_time1+X_time2+X_time3+X_time4+X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)
summary(result)
sum(result$residuals^2)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-16)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))


# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# ACF
acf(data$Value)
acf(result$residuals)

# plot
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-
120000000,400000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)
axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)))

par(new = TRUE)
```

```r
plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-
200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)), lab = c(seq(-
150000000,450000000,by=100000000)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Clean Trend', ylim = c(-
100000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes =
F)
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-100000000,500000000,by=100000000)), lab = c(seq(-
100000000,500000000,by=100000000)))
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd =
c(1.5,1,0.7))
box()
```

```r
# Problem_3
X_time1 = data$index
X_time2 = X_time1**2
X_time3 = X_time1**3
X_time4 = X_time1**4
X1 = data$`1`
X3 = data$`3`
X4 = data$`4`
X5 = data$`5`
X6 = data$`6`
X7 = data$`7`
X8 = data$`8`
X9 = data$`9`
X10 = data$`10`
X11 = data$`11`
X12 = data$`12`
Y = data$Value
result =
lm(Y~0+X_time1+X_time2+X_time3+X_time4+X1+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)
summary(result)
sum(result$residuals^2)
MSE = sum(result$residuals^2)/row
SE_2 = sum(result$residuals^2)/(row-16)
SE = sqrt(SE_2)
logLik(result)
AIC(result)
AIC(result, k=log(row))
```

```r
# 0.95 t_text
t_value_up = qt(0.95, row - 1, lower.tail = T)
t_value_down = qt(0.05, row - 1, lower.tail = T)
up = SE * t_value_up
down = SE * t_value_down

# ACF
acf(data$Value)
acf(result$residuals)

# plot
plot.new()
plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-
200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)
par(new = TRUE)
plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-
200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)
axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)), lab = c(seq(-
150000000,450000000,by=100000000)))

par(new=TRUE)
plot(x = data$index, y = result$residuals,  main = 'Clean Trend', ylim = c(-
100000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes =
F)
abline(h = 0, lwd = 2)
abline(h = up, lty= 2)
abline(h = down, lty= 2)
axis(side = 4, at = c(seq(-100000000,500000000,by=100000000)), lab = c(seq(-
100000000,500000000,by=100000000)))
axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by =
1)))
legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd =
c(1.5,1,0.7))
box()
```